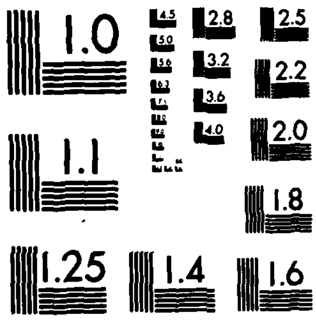
AD-A143 195 BOUNDED-INFLUENCE INFERENCE IN REGRESSION(U) PRINCETON UNIV NJ DEPT OF STATISTICS E RONCHETTI FEB 84 TR-257-SER-2 ARO-19442.17-MA DAAG29-82-K-0178 1/1

UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARO 19442.17-MA	2. GOVT ACCESSION NO. AD-A143 195 N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Technical Report No. 257, Series 2 "Bounded-Influence Inference in Regression"		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Elvezio Ronchetti		8. CONTRACT OR GRANT NUMBER(s) DAAG29-82-K-0178
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Princeton University Princeton, N. J. 08544		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE February 1984 13. NUMBER OF PAGES 39
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE JUL 1 9 1984 S E D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robust testing, influencetion function, likelihood ratio test, F-test, asymptotic power, M-estimators.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two new classes of tests for regression models, likelihood ratio type tests based on quadratic forms of robust estimators, are introduced. Both can be viewed as generalization of the classical F-test. By means of the influence function their robustness properties are investigated and optimally robust tests that maximize the asymptotic power within each class, under the side condition of a bounded influence function, are constructed. Finally, an example based on real data shows that these tests are valuable robust alternatives to the F-test.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

84 07 12 094

Bounded-Influence Inference In Regression

by

Elvezio Ronchetti

Technical Report No. 257, Series 2
Department of Statistics
Princeton University
February 1984

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC COPY INSPECTED 2

This work was supported in part by U.S. Army Research Office Grant Number DAG29-82-K-K0178.

Bounded-Influence Inference
In Regression

by

Elvezio Ronchetti
Department of Statistics
Princeton University

ABSTRACT

Two new classes of tests for regression models, likelihood ratio type tests and tests based on quadratic forms of robust estimators, are introduced. Both can be viewed as generalizations of the classical F-test. By means of the influence function their robustness properties are investigated and optimally robust tests that maximize the asymptotic power within each class, under the side condition of a bounded influence function, are constructed. Finally, an example based on real data shows that these tests are valuable robust alternatives to the F-test.

AMS (1980) SUBJECT CLASSIFICATION: Primary 62F35

Secondary 62J05

KEY WORDS AND PHRASES: robust testing, influence function, likelihood ratio test, F-test, asymptotic power, M-estimators.

1. INTRODUCTION

In this paper we consider the following regression model. Let $\{(x_i, y_i) : i=1,\dots,n\}$ be a sequence of independent identical distributed random variables such that

$$y_i = x_i^T\theta + e_i \qquad i=1,\dots,n \,, \tag{1.1}$$

where y_i is the ith observation, $x_i \varepsilon \mathbb{R}^p$ is the ith row (written as column vector) of the design matrix, $\theta \varepsilon \Theta \subset \mathbb{R}^p$ a p-vector of unknown parameters and $e_i \varepsilon \mathbb{R}$ the ith error. Suppose that e_i is independent of x_i and is distributed according to a normal $N(0,\sigma^2)$. Moreover, denote by $K(x)$ the distribution of the x's and by $F_\theta(x,y)$ the joint distribution of (x_i, y_i) .

Classical estimation and test procedures in regression models are based on the well known method of least squares (LS). This is mostly justified by the Gauss-Markov theorem that states the optimality property of the LS estimator within the class of all linear unbiased estimators. Linearity is a drastic restriction; many maximum likelihood estimators (for example assuming a Cauchy distribution for the errors) are not linear. On the other hand, it is known that the LS estimator is optimal in the class of all unbiased estimators if we assume that the errors are normally distributed. Therefore the restriction to linear estimators can be justified only by normality (or simplicity). But the normal model is never exactly true and in the presence of small departures from the normality assumption on the errors, the LS procedures (estimators and tests) lose efficiency drastically; see Huber (1973), Hampel (1973a, 1978a), Schrader and Hettmansperger (1980),

Ronchetti (1982a,b). Thus, one would prefer to have procedures which are only nearly optimal at the normal model but which behave well in a certain neighborhood of it.

Many robust regression estimators have been proposed in the last years. In section 2 we shall review an importnat class of such estimators, namely the class of M-estimators. Whereas robust estimation theory in regression models has recently received more and more attention (see for instance, Huber (1973), Bickel (1975), Holland and Welsch (1977), Hampel (1978b), Ruppert and Carroll (1980), Krasker and Welsch (1982), Ronchetti and Rousseeuw (1983), Samarov (1983)), the test problem has been somewhat neglected.

From a robustness point of view the classical test procedures based on the LS estimators suffer similar problems as the LS estimators themselves. Although the F-test is moderately robust with respect to the level, it does lose power rapidly in the presence of small departures from the normality assumption on the errors. Recently Schrader and McKean (1977) and Schrader and Hettmansperger (1980) proposed a new class of tests for linear models based on Huber estimates in the full and reduced model, and Sen (1982) found an asymptotic equivalent version of them. Nevertheless, this is only the first step for finding a robust version of the F-test. Like Huber estimators, these tests do not overcome problems caused by situations where the fit is mostly determined by outlying points in the factor space.

The purpose of this paper is twofold. On one side, we introduce new classes of tests that generalize the classical asymptotically equivalent tests, likelihood ratio tests and Wald tests. (The generalization of a third class, $C(\alpha)$ tests, is the subject of a separate paper.) On the other hand,

we propose a solution for the inference problem in regression presenting optimally robust tests that are the natural counterpart of optimally robust estimators and that can be used to construct robust confidence intervals for the parameters.

The paper is organized as follows. Section 2 gives a short overview on bounded-influence estimation in linear models and section 3 presents the approach to robust testing we use to construct a robust version of the F-test. In section 4 and 5 we introduce two new classes of tests, likelihood ratio type tests and tests based on quadratic forms of robust estimators, and we discuss their asymptotic distribution. Each class can be viewed as the natural generalization of a corresponding classical test. In each case the robustness requirement as specified in section 3 leads to an optimally robust test procedure which is a valuable robust alternative to the classical one. Finally, in section 6 we illustrate the excellent performance of optimally robust tests by means of an example based on real data.

2. BOUNDED-INFLUENCE ESTIMATORS

In this section we summarize briefly the results on bounded-influence estimation in regression models. Consider the model (1.1).

One way to cope with the problem of nonrobustness of least squares estimators is to study a large class of estimators generalizing LS, and to select more robust procedures in that class. It appears that M-estimators are most appropriate for this purpose. Suppose for simplicity $\sigma=1$. An M-estimator T_n for the parameter θ is defined as the solution of the implicit equation

$$\sum_{i=1}^{n} \Psi(x_i, y_i; T_n) = 0 , \tag{2.1}$$

for a suitable class of vectorvalued functions $\Psi : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$. Because of the invariance properties of the regression model, an important role is played by the following special class

$$\Psi(x,y;\theta) = \eta(x,y-x^T\theta)x , \tag{2.2}$$

where $\eta : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}$.

There have been several proposals for choosing η . For a stimulating discussion we refer to the papers by Krasker and Welsch (1982) and Huber (1983). Some choices of η are of the form

$$\eta(x,r) = w(x) \cdot \psi(r \cdot v(x)) , \tag{2.3}$$

where $\psi : \mathbb{R} \to \mathbb{R}$, and $w : \mathbb{R}^p \to \mathbb{R}^+$, $v : \mathbb{R}^p \to \mathbb{R}^+$ (weight functions). Huber (1973) uses $w(x) \equiv 1, v(x) \equiv 1$, and Mallows' and Andrews' proposals set $v(x) \equiv 1$ and $w(x) \equiv 1$, respectively. Hill and Ryan proposed $v(x) = w(x)$ and finally, Schweppe suggested choosing $v(x) = 1/w(x)$; see Hill (1977), Krasker and Welsch (1982).

Two tools have been used successfully to study the robustness properties of estimators. The first one, the influence function, was introduced by Hampel (1974) and is essentially the first derivative of an estimator viewed as functional. It describes the normalized influence of an infinitesimal observation on the estimator. The formal definition is the following. Suppose the estimator T_n can be expressed as functional T of the empirical distribution function $F^{(n)}$, $T_n = T(F^{(n)})$. Then the influence function of T at F is given by

$$IF(x,y;T,F) = \lim_{\varepsilon \to 0}[T((1-\varepsilon)F + \varepsilon\delta_{(x,y)}) - T(F)]/\varepsilon \ , \qquad (2.4)$$

where $\delta_{(x,y)}$ is the distribution that puts mass 1 at the point (x,y).

The sencond tool is the change-of-variance function; see Hampel, Rousseeuw, Ronchetti (1981), Ronchetti and Rousseeuw (1983). It can be viewed as the derivative of the asymptotic covariance matrix of the estimator, and describes its infinitesimal stability. From a robustness point of view, a desirable property of these functions is boundedness (in some norm). This means that any (infinitesimal) observation (and therefore any outlier) has a bounded influence on the estimator and on its asymptotic covariance matrix, respectively. Existence conditions and mathematical properties of derivatives of functionals including the influence function are discussed extensively in Clarke (1983) and Fernholz (1983).

The influence function of an estimator defined by (2.2) is given by (see Theorem 4.1 below

$$IF(x,y;\eta,F_\theta) = \eta(x,y-x^T\theta)M^{-1}x \ , \tag{2.5}$$

where $M = \int \eta'(x,r)xx^T d\Phi(r)dK(x)$. By suitable choice of η we can force $\| IF(x,y;\eta,F_\theta)\| < \infty$, for all x and y . Note that $\eta_{LS}(x,r) = r$ defines the LS estimator and

$$IF(x,y;\eta_{LS},F_\theta) = (y-x^T\theta) \cdot (\int uu^T dK(u))^{-1}x$$

is unbounded in x and y .

Hampel's optimality criterion is to put a bound on the IF (measured in some norm) and, under this condition, to minimize the trace of the asymptotic covariance matrix of the estimator at the model. The first condition ensures robustness to the estimator, while the second one is an efficiency condition. If the IF is measured by the Euclidean norm, it turns out that the optimally robust estimator within the class (2.2) is the Hampel-Krasker estimator which is defined by a η-function of Schweppe's form

$$\eta_{HK}(x,r) = \| Ax^{-1}\| \cdot \psi(r \cdot \| Ax\|) \ , \tag{2.6}$$

where $\psi_c(t) = \min(c,\max(t,-c))$ rst the Huber ψ-function, the matrix A is defined implicitly by

$$A^{-1} = E[2\Phi(c/\| Ax\|)-1)xx^T] \tag{2.7}$$

and c is a positive constant depending on the bound on $\|IF\|$; see Hampel (1978b), Krasker (1980).

For an approach to robust statistics using influence functions, we refer to Hampel, Ronchetti, Rousseeuw, Stahel (1984). A critical discussion of bounded-influence estimation in regression can be found in Huber (1983).

3. THE INFINITESIMAL APPROACH TO TESTING

The infinitesimal approach to testing is based on the central notion of influence function. The extension of this concept to tests has been studied by Rousseeuw and the author; see, Ronchetti (1979, 1982a,b), Rousseeuw and Ronchetti (1979, 1981). It turns out that the influence function defined on the test statistic (that is using (2.4) with T=test statistic) is proportional to the influence of an infinitesimal observation on the level and on the power of the test. Therefore, a test statistic with a bounded influence function ensures stability of the level and of the power of the test and guarantees robustness of validity and robustness of efficiency.

Independently, Lambert (1981) introduced in 1979 an influence function for the P-value of a test. For an unconditional test this function is proportional to the influence function of the test statistic. Therefore both functions have the same qualitative behaviour, as far as boundedness and continuity properties are concerned; cf. Lambert (1981).

Hampel's optimality criterion can be extended to tests as follows.

> Find a test which maximizes the asymptotic power within a certain class, under the side condition of a bound on the influence function of the test statistic.

As in the case of estimators, the first condition guarantees robustness and the second one efficiency of the test. We shall use this criterion to select the optimally robust test procedure within the classes of tests defined in the next sections.

For a comparison between different influence functions for tests, see Ronchetti (1982), Field and Ronchetti (1983). The infinitesimal approach is also discussed extensively in Hampel, Ronchetti, Rousseeuw, Stahel (1984).

4. LIKELIHOOD RATIO-TYPE TESTS

Consider the linear model (1.1) and suppose we want to test the linear hypothesis

$$\ell_j(\theta) = 0 \quad , \quad j = q+1,\dots,p \ , \tag{4.1}$$

where $\ell_{q+1}, \dots, \ell_p$ are independent and $0 < q < p$. Through a transformation of the parameter space we can reduce this hypothesis to

$$H_0 : \theta^{(q+1)} = \dots = \theta^{(p)} = 0 \ , \tag{4.2}$$

where $\theta^{(j)}$ denotes the j^{th} component of the vector θ . Let Θ_ω be the subspace of Θ obtained imposing the condition H_0 . The classical test for testing H_0 is the F-test which is equivalent to the likelihood ratio test. It rejects H_0 for "large" values of the test statistic

$$F_n = \sum_{i=1}^{n} [((y_i - x_i^T(T_\omega)_n)/\hat{\sigma})^2 - ((y_i - x_i^T(T_\Omega)_n)/\hat{\sigma})^2]/(p-q) \ , \tag{4.3}$$

where $(T_\Omega)_n$ and $(T_\omega)_n$ are the LS estimates of θ in the full (Θ) and reduced model (Θ_ω) , respectively, and

$$\hat{\sigma}^2 = \sum_{i=1}^{n} (y_i - x_i^T(T_\Omega)_n)^2/(n-p)$$

is the LS (unbiased) estimate of σ^2 .

The aim of this section is to define a class of tests that can be viewed as an extension of the log-likelihood ratio test and therefore of the F-test for linear models.

Let us first introduce the class of functions

$$\tau : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^+ \;,\; (x,r) \to \tau(x,r)$$

with the following properties:

(4. TAU) $\tau(x,r) \not\equiv 0$, $\tau(x,r) \geq 0$ for all $x \in \mathbb{R}^p$, $r \in \mathbb{R}$ and $\tau(x,0) = 0$ for all $x \in \mathbb{R}^p$. $\tau(x,\cdot)$ is differentiable for all $x \in \mathbb{R}^p$. Let $\eta(x,r) := (\partial/\partial r)\,\tau(x,r)$.

(4. ETA1) Assume:

(i) $\eta(x,\cdot)$ is continuous and odd for all $x \in \mathbb{R}^p$,

(ii) $\eta(x,r) \geq 0$ for all $x \in \mathbb{R}^p$ and $r \in \mathbb{R}^+$.

(4. ETA2) $\eta(x,\cdot)$ is differentiable on $\mathbb{R}\setminus\mathcal{D}(x;\eta)$ for all $x \in \mathbb{R}^p$ where $\mathcal{D}(x;\eta)$ is a finite set.

Let $\eta'(x,r) := (\partial/\partial r)\eta(x,r)$ if $x \in \mathbb{R}^p$, $r \in \mathbb{R}\setminus\mathcal{D}(x;\eta)$

$:= 0$ otherwise, and assume

$\sup_r|\eta'(x,r)| < \infty$ for all $x \in \mathbb{R}^p$.

We shall also assume the following regularity conditions:

(4. ETA3) (i) $M := E\eta'(x,r)xx^T$ exists and is nonsingular

and

(ii) $Q := E\eta^2(x,r)xx^T$ exists and is nonsingular.

Definition 4.1 The class of tests $\{\tau\}$ is given by test statistics of the form

$$S_n^2(x_1,\ldots,x_n;y_1,\ldots,y_n) := \\ 2\cdot(p-q)^{-1}n^{-1}\sum_{i=1}^{n}(\tau(x_i,r_{\omega i})-\tau(x_i,r_{\Omega i})) , \tag{4.4}$$

where τ satisfies the conditions (4.TAU), (4.ETA1), (4.ETA2), (4.ETA3),

$$r_{\omega i} := (y_i-x_i^T(T_\omega)_n)/\sigma \quad , \; r_{\Omega i} := (y_i-x_i^T(T_\Omega)_n)/\sigma ,$$

and $(T_\omega)_n$, $(T_\Omega)_n$ are the M-estimators in the reduced and full model, that is

$$\Gamma((T_\omega)_n) = \min\{\Gamma(\theta)|\theta\varepsilon\Theta_\omega\} \tag{4.5}$$

$$\Gamma((T_\Omega)_n) = \min\{\Gamma(\theta)|\theta\varepsilon\Theta\} \tag{4.6}$$

with

$$\Gamma(\theta) := \sum_{i=1}^{n} \tau(x_i,(y_i-x_i^T\theta)/\sigma) . \tag{4.7}$$

"Large" values of S_n^2 are significant.

(In order to give a critical region we shall give the asymptotic distribution of S_n^2 under H_0, see below.)

$(T_\omega)_n$ and $(T_\Omega)_n$ fulfil the equations

$$\sum_{i=1}^{n} \eta(x_i, r_{\omega i})\tilde{x}_i = 0 \tag{4.8}$$

$$\sum_{i=1}^{n} \eta(x_i, r_{\Omega i})x_i = 0 \ . \tag{4.9}$$

Note that $\tilde{x} := (x^{(1)}, \ldots, x^{(q)}, 0, \ldots, 0)^T$ and $(T_\omega)_n = ((T_\omega)_n^{(1)}, \ldots, (T_\omega)_n^{(q)}, 0, \ldots, 0)^T$ (under H_0 the last (p-q) components of θ equal 0 !).

Examples. Define the following functions

$w : \mathbb{R}^p \to \mathbb{R}^+$

$\rho : \mathbb{R} \to \mathbb{R}^+$; $\psi : \mathbb{R} \to \mathbb{R}$, $r \to \psi(r) := (\partial/\partial r)\rho(r)$.

Some choices of τ are of the form $\tau(x,r) = \tilde{w}(x)\rho(r \cdot v(x))$ for certain functions $\tilde{w}(x)$ and $v(x)$. They correspond to the estimators given in section 2.

$\tau(x,r)$	$\eta(x,r)$	estimator corres. to η
$r^2/2$	r	least squares
$\rho(r)$	$\psi(r)$	Huber
$w(x)\cdot\rho(r)$	$w(x)\cdot\psi(r)$	Mallows
$w(x)\cdot\rho(r/w(x))$	$\psi(r/w(x))$	Andrews
$\rho(r\cdot w(x))$	$w(x)\psi(r\cdot w(x))$	Hill and Ryan
$w^2(x)\cdot\rho(r/w(x))$	$w(x)\psi(r/w(x))$	Schweppe

<u>Remark 1</u> In practice, one usually has to estimate the scale parameter σ. A suitable way is to estimate σ in the full model, taking the median absolute deviation or the Proposal 2 estimate of Huber; see Hampel (1974), Huber (1981), p. 137. More precisely, for a given real function χ, one has to solve (4.8), (4.9) and

$$\sum_{i=1}^{n} \chi(r_{\Omega i}) = 0 \tag{4.10}$$

with respect to $(T_\Omega)_n$, $(T_\omega)_n$ and σ. Since we are interested in the robustness properties of these tests, let us compute the influence function of the test statistic S_n.

From now on we put for simplicity $\sigma = 1$.

Let S, T_ω and T_Ω be the functionals corresponding to S_n, $(T_\omega)_n$ and $(T_\Omega)_n$ (see Definition 4.1), that is

$$S(F) = \{2(p-q)^{-1} \int [\tau(u, v-u^T T_\omega(F)) - \tau(u, v-u^T T_\Omega(F))] dF(u,v)\}^{1/2} \tag{4.11}$$

where F is an arbitrary distribution function on $\mathbb{R}^p \times \mathbb{R}$ and T_ω, T_Ω fulfil the system of equations

$$\int \eta(u, v-\tilde{u}^T T_\omega(F))\, \tilde{u}\, dF(u,v) = 0$$

$$\int \eta(u, v-u^T T_\Omega(F))\, u\, dF(u,v) = 0 .$$

(Note that $T_\omega^{(j)}(F) = 0$, for $j = q+1, \ldots, p$ and for all F, and $S_n = S(F_n)$, $(T_\omega)_n = T_\omega(F_n)$, $(T_\Omega)_n = T_\Omega(F_n)$, where F_n is the empirical distribution function of (x_i, y_i) $i=1, \ldots, n$.) The next proposition

gives the influence functions of T_ω, T_Ω and S at the null hypothesis. (Note that, under the null hypothesis, $\theta = \tilde{\theta} = (\theta^{(1)},\ldots,\theta^{(q)},0,\ldots,0)^T$, so $F_{\tilde{\theta}}$ is the model distribution under the null hypothesis.)

<u>Theorem 4.1</u> Assume (4.TAU), (4.ETA1), (4.ETA2) and the following conditions

(4.IF1) $h(\alpha) := \int \eta(x,y-x^T\alpha)x\,dF_{\tilde{\theta}}(x,y)$ exists for all $\alpha \in \Theta \subset \mathbb{R}^p$,
$(\partial/\partial\alpha)h(\alpha)$ exists and is continuous,

(4.IF2) $h(\tilde{\theta}) = 0$.

Then, the influence functions of T_ω, T_Ω and S exist and equal

(i) $IF(x,y;T_\omega,F_{\tilde{\theta}}) = \eta(x,y-x^T\tilde{\theta})\cdot(\tilde{M})^+x$,

(ii) $IF(x,y;T_\Omega,F_{\tilde{\theta}}) = \eta(x,y-x^T\tilde{\theta})\,M^{-1}x$,

(iii) $IF(X,y;S,F_{\tilde{\theta}}) = |\eta(x,y-x^T\tilde{\theta})|\cdot\{[x^T(M^{-1}-(M)^+)x]/(p-q)\}^{1/2}$,

where

$$\tilde{M} := \begin{bmatrix} M_{(11)} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and } (\tilde{M})^+ \text{ denotes the pseudoinverse of } \tilde{M}$$

$$(\tilde{M})^+ = \begin{bmatrix} M_{(11)}^{-1} & 0 \\ 0 & 0 \end{bmatrix} .$$

<u>Proof.</u> Assertions (i) and (ii) follow from Theorem G11.1 of Stahel (1981, p. 116), with $P = F_{\tilde{\theta}}$. His conditions "a", "b", "c" follow from (4.IF1), (4.IF2) and condition "d" from (4.ETA3) (i). Finally, condition "e" follows from (4.ETA2), since $\{(x,y) \mid y-x^T\theta \in \mathcal{D}(x;\eta)\}$ is a regular hyperplane in his sense (Stahel, 1981, p. 12). To show (iii), denote by $\delta_{(x,y)}$ the distribution on $\mathbb{R}^p \times \mathbb{R}$ that puts mass 1 at (x,y) and define the following ε-contaminated distribution

$$F_{\tilde{\theta};\varepsilon,(x,y)} := (1-\varepsilon)F_{\tilde{\theta}} + \varepsilon\delta_{(x,y)} \ .$$

After a straightforward computation we get

$$S^2(F_{\tilde{\theta}}) = 0 \quad \text{(by (4.IF2))} \ ,$$

$$[(\partial/\partial\varepsilon)S^2(F_{\tilde{\theta};\varepsilon,(x,y)})]_{\varepsilon=0} = 0$$

and

$$\begin{aligned}
&[(\partial^2/\partial\varepsilon^2)S^2(F_{\tilde{\theta};\varepsilon,(x,y)})]_{\varepsilon=0} = \\
&\quad 2(p-q)^{-1}[-\eta(x,y-x^T\tilde{\theta})\tilde{x}^T \cdot IF(x,y;T_\omega,F_{\tilde{\theta}}) \\
&\qquad\qquad + \eta(x,y-x^T\tilde{\theta})x^T \cdot IF(x,y;T_\Omega,F_{\tilde{\theta}})] = \\
&\quad 2(p-q)^{-1} \cdot \eta^2(x,y-x^T\tilde{\theta}) \cdot (x^T M^{-1} x - x^T(\tilde{M})^+ x) \ .
\end{aligned}$$

Using L'Hôpital's rule twice, we obtain

$$\begin{aligned}
IF(x,y;S,F_{\tilde{\theta}}) &= \lim_{\varepsilon\to 0}(S(F_{\tilde{\theta};\varepsilon,(x,y)}) - S(F_{\tilde{\theta}}))/\varepsilon \\
&= (\lim_{\varepsilon\to 0} S^2(F_{\tilde{\theta};\varepsilon,(x,y)})/\varepsilon^2)^{1/2} \\
&= (\tfrac{1}{2} \cdot [(\partial^2/\partial\varepsilon^2)S^2(F_{\tilde{\theta};\varepsilon,(x,y)})]_{\varepsilon=0})^{1/2} \ .
\end{aligned}$$

This completes the proof.

As we can see from Theorem 4.1, the influence function corresponding to the F-statistic $(\eta(x,r)=r)$ is unbounded in x and $r = y-x^T\theta$. Our goal is to select in the class $\{\tau\}$ a test with a bounded influence function and as efficient as possible; cf. section 3. To accomplish this we need the distribution of the test statistic defining a τ-test.

Denote by $((T_\Omega)_n)_{(2)}$ the (p-q) vector of the last (p-q) components of $(T_\Omega)_n$ and let $M_{22.1} := M_{(22)} - M_{(21)}M_{(11)}^{-1}M_{(12)}$, where $M_{(ij)}$ are the submatrices of M corresponding to a $q \times (p-q)$ partition of M. Moreover, define $V_n(\theta) := n^{-\frac{1}{2}}\sum_{i=1}^{n}\eta(x_i,y_i-x_i^T\theta)x_i$. Then under the given conditions it is possible to show (see Ronchetti (1982a,b)) that the statistics

$$nS_n^2 , \tag{4.12}$$

$$(p-q)^{-1}V_n^T(\tilde{\theta})(M^{-1}-(\tilde{M})^+)V_n(\tilde{\theta}) \quad \text{and} \tag{4.13}$$

$$W_n^2 := (p-q)^{-1}n((T_\Omega)_n^T)_{(2)}M_{22.1}((T_\Omega)_n)_{(2)} \tag{4.14}$$

have the same asymptotic distribution. More precisely, under the sequence of alternatives

$$H_n : \theta^{(j)} = n^{-\frac{1}{2}}\Delta^{(j)} \quad , \quad j = q+1 ,\ldots, p , \tag{4.15}$$

these statistics are asymptotically distributed as

$$\sum_{j=q+1}^{p}(\lambda_j^{\frac{1}{2}}N_j + (C^T\Delta_{(2)})^{(j)})^2 , \tag{4.16}$$

where $\Delta = (\Delta^{(1)}, \ldots, \Delta^{(p)})$, $N_{q+1}, \ldots, N_p$ are independent univariate standard normal variables, $\lambda_{q+1} \geq \ldots \geq \lambda_p > 0$ are the (p-q) positive eigenvalues of $Q(M^{-1}-(\tilde{M})^{+})$ and C is the Choleski root of $M_{22.1}$ defined by

$$CC^T = M_{22.1} \tag{4.17}$$

$$C^T(M^{-1}QM^{-1})_{(22)}C = \Lambda_{(22)} = \mathrm{diag}(\lambda_{q+1}, \ldots, \lambda_p) . \tag{4.18}$$

Remark 2. A related result on the distribution of the likelihood ratio test statistic when the data do not come from the parametric model under consideration was obtained by Kent (1982); cf. also Foutz and Srivastava (1975).

Remark 3. Under the null hypothesis, (4.16) becomes $\sum_{j=q+1}^{p} \lambda_j (\chi_1^2)_j$, where $(\chi_1^2)_j$ are independent χ^2- random variables with 1 degree of freedom.

We are now ready to solve the optimality problem for infinitesimal robustness of tests (see section 3). We find a τ-test which maximizes the asymptotic power, subject to a bound on the influence function of the test statistic at the null hypothesis. We give the solution to this problem for Huber's regression (that is assuming $\tau(x,r) = \rho(r)$) and in the general case.

Theorem 4.2 Assume: $\tau(x,r)=\rho(r)$. Then the test that maximizes the asymptotic power, under the side condition of a bounded influence of the residual

$$\sup_r IR(r;S,\Phi) \leq b ,$$

is given by Huber's ρ-function

$$\rho_c(r) = r^2/2 \quad \text{if } |r| \leq 0$$
$$= c|r| - c^2/2 \quad \text{otherwise} .$$

<u>Proof.</u> Under the assumption, nS_n^2 is asymptotically distributed as $\lambda \cdot \chi^2_{p-q}(\delta^2)$, where $\lambda = E_\Phi \psi^2 / E_\Phi \psi'$, $\psi(r) = d\rho/dr$ and

$$\delta^2 = [(E_\Phi \psi')^2 / E_\Phi \psi^2] \, \Delta^T_{(2)} (Exx^T)_{(22.1)} \Delta_{(2)} .$$

Therefore the asymptotic power is a monotone increasing function of δ^2. Moreover, the influence function can be factorized in two components, the first one depending only on $r = y - x^T\theta$ (influence of the residual) and the second one depending only in x (influence of position in factor space). Since the first factor equals $|\psi(r)|/E_\Phi \psi'$, the problem we have to solve is equivalent to minimize $E\psi^2/(E\psi')^2$, under a bound on $|\psi(r)|/E\psi'$. But then, using Hampel's Lemma 5 (see Hampel, 1974) we find the solution

$$\psi(r) = \psi_c(r) = r \quad \text{if } |r| < c$$
$$= c \cdot \text{sign}(r) \quad \text{otherwise,}$$

and this proves the theorem.

This class of tests was proposed by Schrader and McKean (1977) and Schrader and Hettmansperger (1980) and carries out in a natural way M-estimation. However, if we look at the influence function of the ρ_c-test, we see that while the factor depending on r is bounded, the second factor depending on x tends to ∞ as $\|x\| \to \infty$. Therefore, the <u>total influence</u> is unbounded and this test suffers the same problems as Huber's estimator when there are outliers in the factor space. This justifies the consideration of the more general class of τ-tests.

The ρ_c-test can be viewed as a likelihood ratio test when the error distribution has a density proportional to $\exp(-\rho_c(r))$. This distribution minimizes the Fisher-information within the gross-error model ("least favorable distribution"). Note that a test of the same type (a likelihood

ratio test under a least favorable distribution) is used by Carroll (1980) who proposes a robust method for testing transformations to achieve approximate normality.

Note that the ρ_c-test is asymptotically equivalent to a test proposed by Bickel (1976) who applies the classical F-test to transformed observations; see Schrader and Hettmansperger (1980) and Huber (1981), p. 197.

In order to state the general optimality result, we first need the following Lemma.

Lemma. Let $c > 0$. If Exx^T is nonsingular, then

(i) for sufficiently large $c > 0$ there exists a symmetric and positive definite $(p \times p)$ matrix $M_S(c,p-q)$ which satisfies the equation

$$E((2\Phi(c\cdot(p-q)^{\frac{1}{2}}/|x^T(M^{-1}-(\tilde{M})^{+})x|^{\frac{1}{2}})-1)xx^T) = M \; ; \qquad (4.19)$$

(ii) M_S converges to Exx^T , when $c \to \infty$;

(iii) Denote by U_S the lower traingular matrix with positive diagonal elements such that $U_S \cdot U_S^T = M_S$ and define

$$\eta_S(x,r) := (||z_{(2)}||/(p-q)^{1/2})^{-1} \cdot \psi_c(r||z_{(2)}||/(p-q)^{1/2}),$$

with $z = U_S^{-1}x$.

Then, $M_S = E\eta_S'(x,r)xx^T$.

Proof. Assertions (i) and (ii) can be shown using the same techniques as in Krasker (1981, Proposition 1), noting that

$$\|\tilde{M}\| < \|M\| \text{ , where } \|M\| := \Sigma_{i,j} m_{ij}^2 \text{ .}$$

(iii) follows using the Choleski decomposition of M and (4.19).

Remark 4. The subscript S for M_S indicates that M_S is the matrix M corresponding to $\eta_S(x,r)$; this function is of the Schweppe form (see section 2).

Theorem 4.3. Assume either (i) $q=p-1$ or (ii) the density of x is spherically symmetric with respect to $x_{(2)} = (x^{(q+1)}, \dots, x^{(p)})$.. Then, the test that solves the optimality problem for infinitesimal robustness within the class of τ-tests, is defined by a function of the form

$$\tau_S(x,r) = (\|z_{(2)}\|/(p-q)^{\frac{1}{2}})^{-2} \cdot {}_c(r\|z_{(2)}\|/(p-q)^{\frac{1}{2}})$$

$$= \rho_{\bar{c}(x)}(r) \text{ .}$$

where $\bar{c}(x) := c\cdot(p-q)^{\frac{1}{2}}/\|z_{(2)}\|$, $z = U_S^{-1}x$ and U_S is the lower triangular matrix (which exists because of the Lemma)

$$E((2\Phi(c\cdot(p-q)^{\frac{1}{2}}/\|z_{(2)}\|)-1)zz^T) = I \text{ .} \tag{4.20}$$

Proof. We show the assertion under condition (i). The proof is similar under (ii). Put

$$M(\eta) = E\eta'(x,r)xx^T \ , \ Q(\eta) = E\eta^2(x,r)xx^T$$

and let $\lambda_p(\eta)$ be the positive eigenvalue of

$$Q(\eta)\cdot(M^{-1}(\eta) - (\tilde{M})^+(\eta)) \ .$$

Moreover, denote by $U(\eta)$ the Choleski decomposition of $M(\eta) = U(\eta)\cdot U^T(\eta)$, here $U(\eta)$ is a lower traingular matrix with positive diagonal elements. We have to solve the following problem. For a given $b > 0$, find a test which maximizes $M_{(22.1)}/\lambda_p$, under the side condition

$$\sup_{x,r}(|\eta(x,r)|/u_{pp})\cdot(x^T(M^{-1}-(\tilde{M})^+)x)^{\frac{1}{2}} \leq b \ . \tag{4.21}$$

Since $U^T(M^{-1}-(\tilde{M})^+)U = I-\tilde{I}$ we obtain

$$\lambda_p(\eta) = (U^{-1}Q\,U^{-T})_{pp} = E\eta^2(x,r)|(U^{-1}x)^{(p)}|^2 \ , \tag{4.22}$$

$$M_{(22.1)}(\eta) = (u_{pp}(\eta))^2 \ . \tag{4.23}$$

Moreover, (4.21) becomes

$$\sup_{x,r}(|\eta(x,r)|/u_{pp})\cdot|(U^{-1}x)^{(p)}| \leq b \ . \tag{4.24}$$

Choose $c > 0$ such that $b = c/(U_S)_{pp}$, where U_S is defined by Lemma (iii), (this c exists because $c^2/U_S)^2_{pp} \geq c^2/\mathrm{tr}\, M_S \geq c^2/E\|x\|^2 \to \infty$, when $c \to \infty$) and assume, without loss of generality,

$$u_{pp}(\eta) = (U_S)_{pp} . \tag{4.25}$$

(The multiplication of the test statistic by a positive constant does not change the test!) Combining (4.22), (4.23), (4.24) and (4.25), the original problem reduces to minimize $E\eta^2(x,r)\cdot|(U^{-1}x)^{(p)}|^2$, under the conditions (4.25) and

$$\sup_{x,r}|\eta(x,r)|\cdot|(U^{-1}x)^{(p)}| \leq c . \tag{4.26}$$

Now,

$$E\eta^2(x,r)\cdot|(U^{-1}x)^{(p)}|^2 = -E\, r^2\cdot|(U_S^{-1}x)^{(p)}|^2 + 2$$
$$+ E(\eta(x,r)\cdot(U^{-1}x)^{(p)} - r\cdot(U_S^{-1}x)^{(p)})^2 ,$$

since

$$E(\eta(x,r)\cdot r\cdot(U^{-1}x)^{(p)}\cdot(U_S^{-1}x)^{(p)})$$
$$= (U^{-1}\cdot E(\eta(x,r)\cdot r\cdot xx^T)\cdot U_S^{-1})_{pp} ,$$

and integrating by parts,

$$= (U^{-1}\cdot E(\eta'(x,r)xx^T)\cdot U_S^{-1})_{pp} = (U^{-1}M\ U_S^{-1})_{pp}$$

$$= (U^{-1}U\ U^T U_S^{-1})_{pp} = (U^T U_S^{-1})_{pp} = u_{pp}/(U_S)_{pp} = 1\ ,$$

where in the last equalities we have used (4.25) and $U_{(12)} = 0$. Thus, minimizing $E\eta^2(x,r)\cdot|(U^{-1}x)^{(p)}|^2$, under the conditions (4.25) and (4.26) is equivalent to minimizing

$$E\{(\eta(x,r)\cdot(U^{-1}x)^{(p)} - r\cdot(U_S^{-1}x)^{(p)})^2\}\ ,$$

subject to (4.26). Clearly, the optimal η^* must satisfy

$$\eta^*(x,r)\cdot((U^{-1}(\eta^*)x)^{(p)}) = \psi_c(r\cdot(U_S^{-1}x)^{(p)}).$$

Therefore,

$$\eta_S(x,r) = |z_{(2)}|^{-1}\cdot\psi_c(r|z_{(2)}|)\ ,$$

where $z = U_S^{-1}x$, solves this extremal problem. Any other solution defines a test which has the same influence function and the same asymptotic power and in this sense is equivalent to η_S . This completes the proof.

Note that the η-functions defining optimally bounded influence tests and optimally bounded influence estimators are of the same form (see

Theorem 4.3 and (2.6)), namely of Schweppe's form. There is a difference only in the weights: the optimal weights for the test take into account that (after standardization) only the last (p-q) components are of interest for the testing problem.

From these optimally robust tests one can derive robust confidence regions for the parameters and a robust version of stepwise regression. These procedures can be easily implemented in a package for robust regression. Especially, it is planed to integrate them in ROBETH, a package of robust linear regression programs which have been written at the ETH Zurich and is still under development; see Marazzi (1980).

5. TESTS BASED ON QUADRATIC FORMS OF ROBUST ESTIMATORS

Let T_n be a general M-estimator defined by (2.1). Huber (1967) shows that $n^{\frac{1}{2}}(T_n-\theta)$ is asymptotically normal with mean 0 and asymptotic covariance matrix

$$V = B^{-1} A B^{-1} , \tag{5.1}$$

where $B = -E\partial\Psi/\partial\theta$, $A = E\Psi\Psi^T$. Partition the matrix V in $q \times q$, $q \times (p-q)$, $(p-q) \times q$, $(p-q) \times (p-q)$ blocks and denote them by $V_{(11)}$, $V_{(12)}$, $V_{(21)}$, $V_{(22)}$, respectively. Moreover denote by $(T_n)_{(1)}$ and $(T_n)_{(2)}$ the vectors with the first q components and the last $(p-q)$ components of T_n , respectively.

<u>Definition 5.1</u> Let T_n be a general M-estimator defined by (2.1). Then the test statistic

$$R_n^2 := (T_n)_{(2)}^T (V_{(22)})^{-1} (T_n)_{(2)} / (p-q)$$

defines a class of tests for testing the hypothesis H_0 (see (4.2)). "Large" values of R_n^2 are significant.

From Definition 5.1 we see that R_n^2 is a quadratic form of the estimator $(T_n)_{(2)}$ with respect to its asymptotic covariance matrix. Therefore, under the conditions of Huber (1967) and (4.15), $(p-q)nR_n^2$ has asymptotically a χ^2-distribution with p-q degrees of freedom and noncentrality parameter

$$\delta^2 = \Delta_{(2)}^T (V_{(22)})^{-1} \Delta_{(2)} . \tag{5.2}$$

Under the conditions of Theorems 4.2 and 4.3 and in view of (4.12) and (4.14), it follows easily that likelihood ratio-type tests and tests based on R_n^2 are asymptotically equivalent. However, at least in some situations, the latter seem to have a more liberal small sample behaviour; cf. Schrader and Hettmansperger (1980).

Remark 1. A class of tests based on quadratic forms of robust estimators was proposed by Stahel (1981) and Samarov and Welsch (1982) in the case $q=p-1$ and for M-estimators of the form (2.2). Schrader and Hettmusperger (1980) consider the same class of tests in the special case $\eta(x,r) = \psi(r)$.

Let us now compute the influence function of this new class of tests. Since the computations are similar to those in Theorem 4.1, we drop them. Note only that we compute the influence function of R and not of R^2, the latter being identical to zero. This does not affect the test since R and R^2 define the same test. The influence function of R at the model $F_{\tilde{\theta}}$ is given by

$$IF(x,y;R,F_{\tilde{\theta}}) = (z^T(V_{(22)})^{-1}z/(p-q))^{\frac{1}{2}} , \tag{5.3}$$

where $z = IF(x,y;T_{(2)},F_{\tilde{\theta}})$ is the influence function of the estimator $T_{(2)}$ at the model $F_{\tilde{\theta}}$.

Thus we have

$$\sup_{x,y}|IF(x,y;R,F_{\tilde{\theta}})| = (p-q)^{-\frac{1}{2}}\gamma_s^*(T_{(2)},F_{\tilde{\theta}}) , \tag{5.4}$$

where γ_s^* is the self-standardized sensitivity of the estimator $T_{(2)}$; c.f. Krasker and Welsch (1982).

Note again that the Wald test, which is defined through the test statistic R_n^2 when T_n is the LS estimator, has an unbounded influence function. Therefore our goal is to find a test based on R_n^2 that maximizes the asymptotic power under a bound on the influence function (5.3). In view of (5.2) and (5.4), this problem is equivalent to the following estimation problem

> Find a function Ψ defining an M-estimator T that minimizes the asymptotic covariance matrix $V_{(22)}$ of $T_{(2)}$, under a bound on the self-standardized sensitivity of $T_{(2)}$.

It turns out that it is impossible to find an M-estimator that minimizes in the strong sense the asymptotic covariance matrix, subject to a bound on the self-standardized sensitivity. A counterexample can be found in Stahel (1981). However, the following admissibility result can be shown.

<u>Definition 5.2</u> A test defined by Ψ^* dominates a test defined by Ψ if as.power$_{\Delta_{(2)}}\Psi^* \geq$ as.power$_{\Delta_{(2)}}\Psi$, and there is a $\bar{\Delta}_{(2)}$ such that as.power$_{\bar{\Delta}_{(2)}}\Psi^* >$ as.power$_{\bar{\Delta}(2)}\Psi$.

<u>Definition 5.3</u> A test defined by Ψ is called admissible if there is no test that dominates it.

<u>Theorem 5.1</u> For a given $c \geq 1$, let $C_c(\Psi)$ be the class of tests given in Definition 5.1 and such that $\sup_{x,y}|IF(x,y;R,F_{\tilde{\theta}})| \leq c$. Then, the test defined by the following function Ψ^* is admissible within $C_c(\Psi)$:

$$\Psi^*_{(1)}(x,y;\theta) = r - z_{(1)} ,$$

$$\Psi^*_{(2)}(x,y;\theta) = (z_{(2)}/\|z_{(2)}\|) - \psi_c(r\|z_{(2)}\|) ,$$

where $\Psi^*_{(1)}$ and $\Psi^*_{(2)}$ denote the first q and last $(p-q)$ components of Ψ^* respectively, $r = y-x^T\theta$, $z = Dx$, and D is defined implicitly by the matrix equation

$$E\Psi(x,y;\theta)rx^T = I .$$

Proof. In veiw of (5.2) and (5.4) the testing problem is equivalent to that of finding an M-estimator T which is admissible with respect to the asymptotic covariance matrix $V_{(22)}$, subject to the condition $(p-q)^{-\frac{1}{2}}\gamma^*_S(T_{(2)},F_{\tilde{\theta}}) \leq c$. This estimation problem is a special case of the problem solved in Stahel (1981, p. 40) (cf. also Hampel, Ronchetti, Rousseeuw, Stahel (1984), Theorem 1, section 4.4) and the result follows.

If we restrict ourselves to the class (2.2), either under the conditions of Theorem 4.2 or Theorem 4.3, it is easy to see that the test defined by Ψ^* is asymptotically equivalent (that is, it has the same asymptotic power and the same influence function) to the corresponding optimally robust tests given in those theorems. Therefore, under those conditions, Ψ^* is in fact storng optimal.

6. EXAMPLE

The following example based on a real data set should give an idea of the small sample behaviour of our optimally robust tests. The data are taken from Draper and Smith (1966, p. 104 ff.). We have the following variables:

Y = response or number of pounds of steam used per month,

X_8 = average atmospheric temperature in the month (in ^0F)

X_6 = number of operating days in the month.

Table 1 shows the data.

We consider the linear model

$$Y = \alpha + \beta_8 X_8 + \beta_6 X_6 + e ,$$

and we want to test the hypothesis

$$H_0 : \beta_6 = 0 .$$

The factor space is given by Figure 1 and the observations are plotted in Figure 2a.

From Figure 1 we see that there exist two outliers in the factor space. We want to study the behaviour of the $\log_{10}$P-values of the F- , ρ- and optimal τ-test when the observation (y_7) corresponding to the point $(X_8 = 74.4$, $X_6 = 11)$ varies between 0 and 20. (Its actual value is 6.36)

The tests under study are defined by the following functions

test	$\tau(x,r)$	
F	$r^2/2$	
ρ_{c_1}	$\rho_{c_1}(r)$	(see Theorem 4.2)
optimal τ	$\rho_{c_2/\lvert z^{(2)}\rvert}(r)$	$(z^{(2)} = (U_S^{-1}x)^{(2)})$ (see Theorem 3)

The scale parameter σ was estimated in the full model using Huber's Proposal 2. Note that in this case the optimal τ-test is equivalent to the corresponding test based on quadratic forms given in section 5. The constants c_1 and c_2 were chosen such that the corresponding tests have a given efficiency, say .95, at the normal model (that is, when x_i and e_i are normally distributed). We obtained the following values:

$$c_1 = 1.345 \qquad c_2 = 2.67 .$$

Figure 2b shows the overall excellent behaviour of the τ-test (strongly significant for all y!), the good behaviour of the ρ_c-test (at least for $y \geq 8$) which is still significant (at the 5% level) in the region $y \geq 8$ and the bad behaviour of the F-test which becomes even not significant for $8.7 \leq y \leq 18.7$!

6. FURTHER RESEARCH DIRECTIONS

Approximate critical regions for optimally robust tests derived in this paper can be obtained using the asymptotic distribution of the test statistic, see (4.16) and (5.2). This approximation can be improved in two ways.

The first possibility is to use S_n^2 or R_n^2 defined in section 4 and 5 respectively, as test statistic for a permutation test. This would guarantee, on one side an exact level α-test (a property of permutation, tests) and on the other side a high robustness of efficiency (a property of S_n^2 and R_n^2) . This idea has been applied for constructing a confirmatory test in connection with a Swiss hail experiment; see Hampel, Schweingruber, Stahel (1982). Some work is needed to justify this combined procedure from a theoretical point of view; cf. Lambert (1982).

The second way is to find better approximations to the distribution of S_n^2 and R_n^2 . A promising approach is small sample asymptotics that has been used successfully in the location case; see Hampel (1973b), Field and Hampel (1982). This is subject of ongoing research.

ACKNOWLEDGMENTS

Part of this work is taken from the author's Ph.D thesis, which was written at the ETH Zurich under the generous guidance of Prof. F.R. Hampel. The author is grateful to I. Olkin, C. Field, P. Rousseeuw, W. Stahel, R. Welsch, A. Samarov, V. Yohai for their valuable suggestions. Partial support of ARO (Durham) contract #DAAG29-82-K-K0178 is also gratefully acknowledged.

REFERENCES

Bickel, P.J. (1975). One-step Huber estimates in the Linear Model. *Journal of the American Statistical Association.* 70 428-434.

Bickel, P.J. (1976). Another Look at Robustness: A Review of Reviews and Some New Development. *Scandinavian Journal of Statistics.* 3 145-168.

Carroll, R.J. (1980). A Robust Method For Testing Transformations to Achieve Approximate Normality. *Journal of the Royal Statistical Society.* Series B 42 71-78.

Clarke, B.R. (1983). Uniqueness and Fréchet Differentiability of of Functional Solutions to Maximum Likelihood Type Equations. *The Annals of Statistics.* 11 1196-1205

Draper, N.R., Smith, H. (1966). *Applied Regression Analysis.* New York: Wiley.

Fernholz, L.T. (1983). *Von Mises Calculus For Statistical Functionals.* Lecture Notes in Statistics. 19 New York: Springer.

Field, C.A., Hampel, F.R. (1982). Small-sample Asymptotic Distributions of M-estimators of Location. *Biometrika.* 69 29-46.

Field, C.A., Ronchetti, E. (1983). A Tail Area Influence Function and Its Application to Testing. Statistics Research Report No. 1. Dept. of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax.

Foutz, R.V., Srivastava, R.C. (1977). The Performance of the Likelihood Ratio Test When the Model is Incorrect. *The Annals of Statistics.* 5 1183-1194.

Hampel, F.R. (1973a). Robust Estimation: A Condensed Partial Survey. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete.* 27 87-104.

Hampel, F.R. (1973b). Some Small-sample Asymptotics. *Proceedings of the Prague Symposium on Asymptotic Statistics.* Prague. 109-126.

Hampel, F.R. (1974). The Influence Curve and Its Role in Robust Estimation. *Journal of the American Statistical Association.* 69 383-393.

Hampel, F.R. (1978a). Modern Trends in the Theory of Robustness. *Mathematische Operationsforschung und Statistik.* Series Statistics. 9 425-442.

Hampel, F.R. (1978b). Optimally Bounding the Gross-error-sensitivity and the Influence of Position in Factor Space. *1978 Proceedings of the ASA Statistical Computing Section.* ASA. 59-64.

Hampel, F.R., Ronchetti, E., Rousseeuw, R.J. Stahel, W. (1984). *Robust Statistics: The Infinitesimal Approach.* New York: Wiley (to appear).

Hampel, F.R., Rousseeuw, P.J., Ronchetti, E. (1981). The Change-of-variance Curve and Optimal Redescending M-estimators. *Journal of the American Statistical Association.* 76 643-648.

Hampel, F.R., Schewingruber, M., Stahel, W. (1983). Das Ergebnis des konfirmatorischen Tests im Grossversuch IV zür Hagelabwehr. Research Report No. 35. Fachgruppe für Statistik, ETH Zürich.

Hill, R.W. (1977). Robust Regression When There Are Outliers in the Carriers. Ph.D. Thesis, Harvard University. Cambridge, Mass.

Holland, P.W., Welsch, R.E. (1977). Robust Regression Using Iteratively Reweighted Least-Squares. *Communications in Statistics.* A6 813-828.

Huber, P.J. (1967). The Behaviour of the Maximum Likelihood Estimates Under Nonstandard Conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: University of California Press. 221-233.

Huber, P.J. (1973). Robust Regression: Asymptotic, Conjectures and Monte Carlo. *The Annals of Statistics.* 1 799-821.

Huber, P.J. (1981). *Robust Statistics.* New York: Wiley.

Huber, P.J. (1983). Minimax Aspects of Bounded-Influence Regression. *Journal of the American Statistical Association.* 78 66-80.

Kent, J.T. (1982). Robust Properties of Likelihood Ratio Tests. *Biometrika.* 69 19-27.

Krasker, W.S. (1980). Estimation in Linear Regression Models With Disparate Data Points. *Econometrica.* 48 1333-1346.

Krasker, W.S., Welsch, R.E. (1982). Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association.* 77 595-605.

Lambert, D. (1981). Influence Functions For Testing. *Journal of the American Statistical Association.* 76 649-657.

Lambert, D. (1982). Robust Permutation Tests. Technical Report No. 234. Department of Statistics, Carnegie-Mellon University. Pittsburgh.

Marazzi, A. (1980). Robust Linear Regression Programs in ROBETH. Research Report No. 23. Fachgruppe für Statistik, ETH Zürich.

Ronchetti, E. (1979). Robustheitseigenschaften von Tests. Diploma Thesis. ETH Zürich.

Ronchetti, E. (1982a). Robust Alternatives to the F-test For the Linear Model. in: Grossmann W., Pflug C., Wertz, W. eds., *Probability and Statistical Inference*. Dortrech: Reidel. 329-342.

Ronchetti, E. (1982b). Robust Testing in Linear Models: The Infinitesimal Approach. Ph.D. Thesis. ETH Zürich.

Ronchetti, E., Rousseeuw, P.J. (1983). The Change-of-Variance Function and Robust Estimators of Regression. *Proceedings of the ASA Statistical Computing Section*. ASA. 165-168.

Rousseeuw, P.J., Ronchetti, E. (1979). The Influence Curve For Tests. Research Report No. 21. Fachgruppe für Statistik, ETH Zürich.

Rousseeuw, P.J., Ronchetti, E. (1981). Influence Curves For General Statistics. *Journal of Computational and Applied Mathematics*. 7 161-166

Ruppert, D., Carroll, R.J. (1980). Trimmed Least Squares Estimation in the Linear Model. *Journal of the American Statistical Association*. 75 828-838.

Samarov, A.M. (1983). Bounded-Influence Regression Via Local Minimax Mean Square Error. Technical Report No. 40. MIT. Cambridge, MA.

Samarov, A.M., Welsch, R.E. (1982). Computational Procedures For Bounded-Influence Regression. *Proceedings in Computational Statistics*, COMPSTAT 1982 (to appear).

Schrader, R.M., Hettmansperger, T.P. (1980). Robust Analysis of Variance Based Upon A Likelihood Ratio Criterion. *Biometrika*. 67 93-101.

Schrader, R.M., McKean, J.W. (1977). Robust Analysis of Variance. *Communications in Statistics*. A6 879-894.

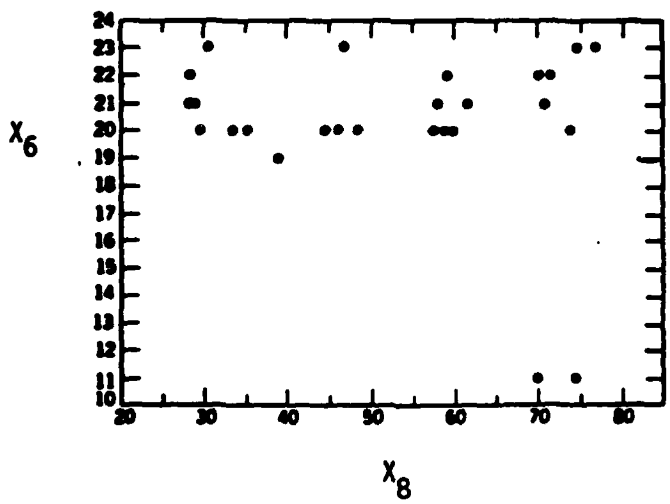
Sen, P.K. (1982). On M-tests in Linear Models. *Biometrika*. 69 245-248.

Stahel, W. (1981). Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen für Kovarianzen. Ph.D. Thesis, ETH Zürich.

Table 1

X_8	X_6	Y
35.3	20	10.98
29.7	20	11.13
30.8	23	12.51
58.8	20	8.40
61.4	21	9.27
71.3	22	8.73
74.4	11	6.36
76.7	23	8.50
70.7	21	7.82
57.5	20	9.14
46.4	20	8.24
28.9	21	12.19
28.1	21	11.88
39.1	19	9.57
46.8	23	10.94
48.5	20	9.58
59.3	22	10.09
70.0	22	8.11
70.0	11	6.83
74.5	23	8.88
72.1	20	7.68
58.1	21	8.47
44.6	20	8.86
33.4	20	10.36
28.6	22	11.08

Figure 1: The variables X_6 , X_8



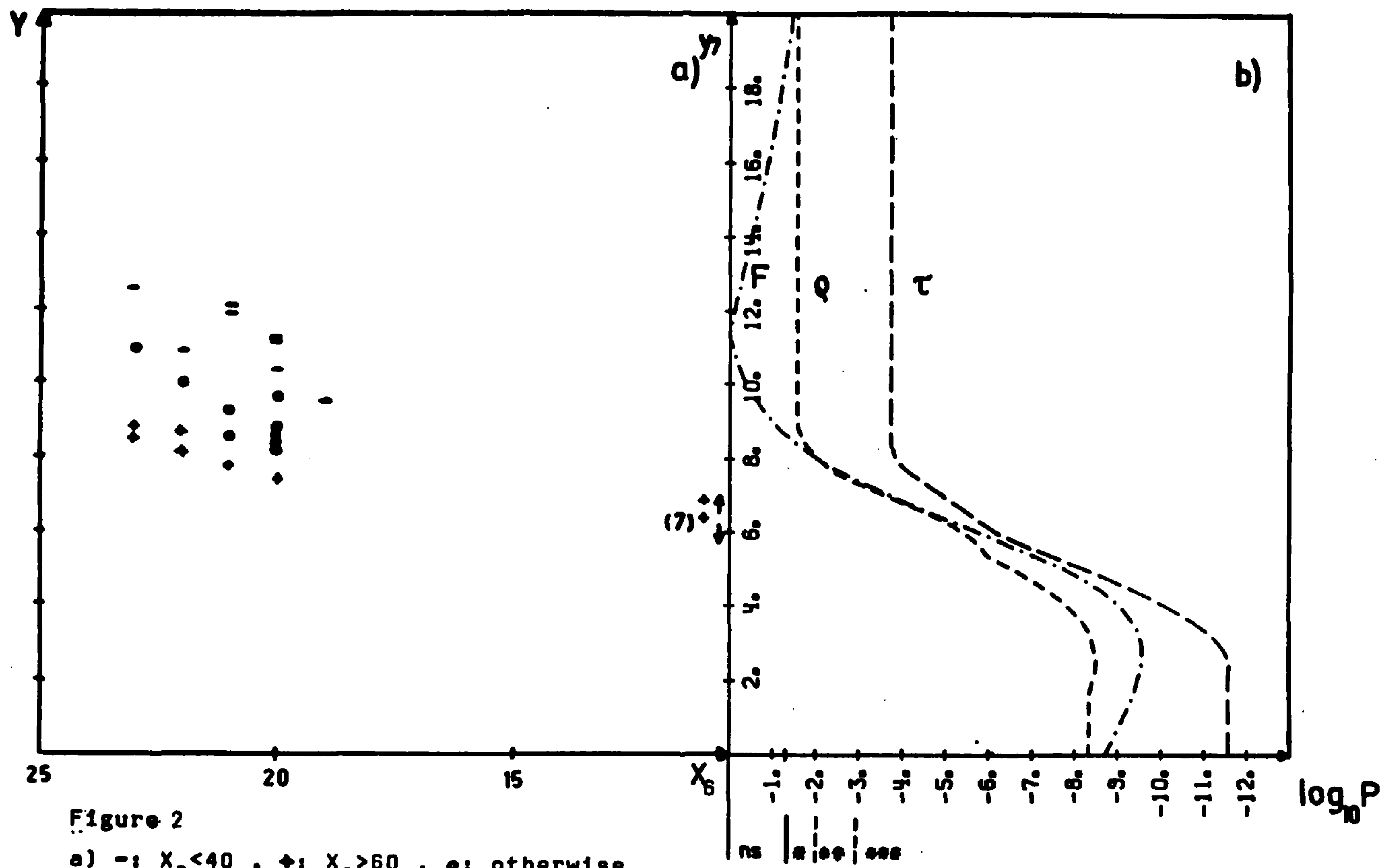


Figure 2

a) -: $X_8<40$, +: $X_8>60$, •: otherwise

b) $\log_{10}$P-values against y_7

END

FILMED

9-84

DTI